Physics-Based Synthetic Data Model for Automated Segmentation in Catalysis Microscopy

Maurits Vuijk, Gianmarco Ducci, Luis Sandoval, Markus Pietsch, Karsten Reuter, Thomas Lunkenbein, Christoph Scheurer

Microscopy and Microanalysis, 2025, **31**, ozae130
https://doi.org/10.1093/mam/ozae130
Advance access publication 13 January 2025
Original Article

Microscopy AND Microanalysis

Physics-Based Synthetic Data Model for Automated Segmentation in Catalysis Microscopy

Maurits Vuijk[1], Gianmarco Ducci[1], Luis Sandoval[1], Markus Pietsch[2], Karsten Reuter[1], Thomas Lunkenbein[1,*], and Christoph Scheurer[1,3,*]

[1]*Fritz-Haber-Institut der Max-Planck-Gesellschaft, Berlin 14195, Germany*
[2]*Technische Universität München, Department of Chemistry, München 80333, Germany*
[3]*Institute of Energy Technologies, Fundamentals of Electrochemistry (IET-1), Forschungszentrum Jülich GmbH, 52425 Jülich, Germany*

***Corresponding author:** Christoph Scheurer, E-mail: scheurer@fhi.mpg.de; Thomas Lunkenbein, E-mail: lunkenbein@fhi-berlin.mpg.de

Abstract

In catalysis research, the amount of microscopy data acquired when imaging dynamic processes is often too much for nonautomated quantitative analysis. Developing machine learned segmentation models is challenged by the requirement of high-quality annotated training data. We thus substitute expert-annotated data with a physics-based sequential synthetic data model. We study environmental scanning electron microscopy (ESEM) data collected from isopropanol oxidation to acetone over cobalt oxide as an example. Upon applying a temperature program during the reaction a phase transition occurs, reducing the catalyst selectivity toward acetone. This is accompanied on the micrometer ESEM scale by the formation of cracks between the pores of the catalyst surface. We aim to generate synthetic data to train a neural network capable of semantic segmentation (pixel-wise labeling) of this ESEM data. This analysis will lead to insights into this phase transition. To generate synthetic data that approximates this transition, our algorithm composes the ESEM images of the room-temperature catalyst with dynamically evolving synthetic cracks satisfying physical construction principles, gathered from qualitative knowledge accessible in the ESEM data. We mimic the surface crack growth propagation along surface paths, avoiding close vicinity to nearby pores. This physics-based approach results in a lowered rate of false positives compared to a random approach.

Key words: computer vision, ESEM, LSTM, machine learning, synthetic data, U-NET

Introduction

Catalyst research frequently involves the use of sequential operando electron microscopy imaging techniques that capture time-series data. These imaging techniques are powerful tools to capture dynamic processes and phase changes at different scales of the catalyst (Uwins et al., 1997). However, handling and analysis of the large amount of data acquired at high frame rates (often over days) during such imaging processes presents a significant challenge for researchers. Thus, there is a growing need for automated segmentation processes that can efficiently analyze these large sets of electron microscopy data. For this purpose, many different computer vision algorithms and neural networks have been developed which enable automatic analyses focusing on different aspects of the collected image data, such as interface boundaries (Schneider et al., 2016) and carbon nanostructures (Toth et al., 2013).

For scanning electron microscopy (SEM) data, reliable semantic segmentation (pixel-wise labeling) is required to extract meaningful statistical data from the image sequences. Manually segmenting such datasets is tedious and not time efficient (Madsen et al., 2018). As a solution, Machine Learning models have been successfully employed for this purpose (Groschner et al., 2021). Besides saving time, using an automated solution also lowers bias and provides reproducibility and self-consistency. However, to train a network capable of reliable semantic segmentation, a large amount of annotated training data is required. Alternatively, unsupervised learning methods can be used, but these methods tend to lead to less reliable results (Ede, 2021).

Automated segmentation processes using Machine Learning models have shown great utility for image segmentation and classification in electron microscopy. A large amount of research has been done for different model architectures and applications within this field. Supervised semantic segmentation models utilizing convolutional neural network (CNN) architectures have since become widespread, with many different architectures and implementations being developed, such as the PixelNet, which segments microstructures in SEM images (DeCost et al., 2019). By convolution of trained filters with input images at different resolutions, CNNs are able to extract feature information between details and domains of images. The power of the CNN has also been combined with recurrent neural network elements (Liu et al., 2021), such as the long short term-memory (LSTM) element (Hochreiter & Schmidhuber, 1997), to allow networks to form connections between images in image stacks. The LSTM cell is well suited for the identification of evolving features in an image sequence, due to its ability to remember relevant information (newly appearing and established features) and forget obsoleted information (fading features).

One widely used model is U-NET, which was originally developed for biomedical applications to segment electron microscopy images of cells. The core idea of the U-NET architecture is an encoder–decoder-like structure with skip

Received: May 8, 2024. **Revised:** August 5, 2024. **Accepted:** December 5, 2024

connections. Each of the contracting units in the encoder is mirrored with an inverse expanding unit in the decoder, linked from the corresponding contracting unit by a skip connection. This leads to the namesake U shape of the network (Ronneberger et al., 2015).

The U-NET architecture has been shown to be versatile and applicable beyond the original application of biomedical imaging. It is well suited for segmenting transmission electron microscopy (TEM) images, where it can exploit the regular structures in the images (Bermudez-Chacon et al., 2018), as well as SEM images of materials and surfaces (Shah et al., 2023).

As an alternative to assembling expert-annotated datasets, synthetic data methods have been developed. Synthetic data methods involve generating data along with ground truth using a model, and then using the generated data to train a neural network for segmentation of the target (real) data. With atomic-resolution microscopy techniques such as TEM, it is possible to use established theoretical models to generate accurate synthetic data (Ziatdinov et al., 2017). However, when looking at lower-magnification SEM images of catalyst surfaces, the often chaotic nature of the macroscopic surface structure (Sandoval-Diaz et al., 2024) demands a more ad-hoc visual approach to synthetic data generation. By examining the important features of the data empirically, it is possible to create a synthetic data model that generates visually similar images that respect the physical behavior of the features. The work of Trampert et al. (2021) shows such results that directly inspired this work. In their work, they use exemplar-based inpainting as a synthetic data generator to train a classification network. To do so, a texture dictionary is built up from real samples to generate new backgrounds. Features are then generated according to predefined parameters and superimposed on these backgrounds to generate the synthetic data. Another approach involving algorithmic features is explored in Govind et al. (2024).

In this work, we build upon this previous work in two ways: We use a physics-based model to generate sequential synthetic data with inpainted evolving features rather than a purely parametric model. The physics-based model naturally generates a distribution of cracks with similar qualities (length, tortuosity) to those found in the real data, and confers an important improvement in accuracy, especially when it comes to false positives, over the parametric model. As our data consist of a time-series in which slow evolution of the features is clearly visible, we introduce a temporal element to the neural network architecture by appending a convolutional LSTM block (Chang, 2019) to the usual U-NET architecture. The convolutional LSTM block allows the neural network to take advantage of relations across the time dimension of the time-series in addition to the spatial dimensions of the images. To train a neural network containing an LSTM block, sequential annotated training data is required. For this reason, the synthetic data generator was modified to generate sequences. In addition, an important advantage of using the U-NET in the context of synthetic data is that the U-NET architecture is very good at learning transformational invariance from augmented datasets (Wilms et al., 2017). As a synthetic data approach could be regarded as an extreme form of data augmentation, as well as being combined with data augmentation of the synthetic dataset, this property of the network architecture is advantageous (Fawaz et al., 2018).

We thus propose an approach to overcome the annotation challenge in training semantic segmentation neural networks. Instead of relying on expert-annotated data, we substitute it with a physics-based sequential synthetic data model. This model mimics the dynamic processes involving the relevant features observed in our example system. We investigated the selective oxidation of isopropanol to acetone over a cobalt oxide catalyst surface, where we observed the formation of crack features during the reaction. By training a U-NET with LSTM block using synthetic data we aim to train a model that can reliably segment this crack feature. This reaction has been previously investigated using Co_3O_4 nanoparticles where it was found that the catalytic performance can be grouped into low- and high-temperature activity regimes (Anke et al., 2019). It was concluded that the reduction of Co^{3+} to Co^{2+} and the poisoning of the active sites due to carbonaceous deposits caused the change in the reactivity.

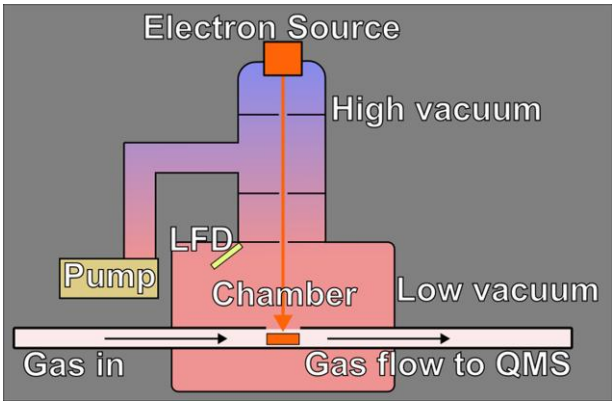

Fig. 1. Experimental ESEM setup. ESEM is a technique that allows for SEM images to be taken under atmospheric conditions (Sandoval-Diaz et al., 2020, 2024). By using pumps to create a pressure gradient, the requirements of high vacuum for the operation of the electron gun, and low vacuum for the gas flow required for catalytic experiments within the sample chamber can be satisfied. LFD is the large field detector within the ESEM, positioned near the chamber aperture. QMS is a quadrupole mass spectrometer, allowing data collection of the output of the experiment.

Materials and Methods

ESEM Experiment

In this section, we introduce the sample and the experiment that was performed to obtain the experimental dataset, and we describe the goals and challenges with regards to the analysis of the dataset. A constant gas flow of isopropanol and oxygen was present during the experiment and the catalyst was heated with staggered heating to 500°C using a laser heating setup. A time-series, consisting of 1,600 frames with a scan time of 35.40 s, of the evolution of the catalyst surface was gathered during the experiment in an operando SEM experiment, as depicted in Figure 1. For further details about the experimental setup, see Supplementary Material S1.

Within the environmental SEM (ESEM), the large field detector (LFD) was used as the electron detector. The LFD collects signals from both secondary and backscattered electrons by proxy through ionization of the gas present in the ESEM chamber. As such, the contrast obtained in this ESEM setup contains elements of Z-contrast and topological contrast (Donald, 2003). Since the gas is directly involved in the measurement, changes to the gas composition and

temperature affect the signal captured by the detector. Temperature changes are especially problematic, as they effectively lead to drifting of the sample (Danilatos, 1983).

The initial surface of the catalyst displays a complex morphology with prominent (bright) ridges and (dark) holes (see Fig. 3 below). After being heated to 400°C, a change in the surface texture occurs that coincides with a large drop in the selectivity toward acetone of the catalyst, visible in Figure 2. On the ridges of the catalyst surface, crack structures start to appear, i.e. the cracks remain spatially separated from the original holes of the catalyst morphology. Following the formation of the crack structures, the surface structure of the catalyst now has two key (dark) features: "holes" and "cracks". The *in situ* formation of cracks is seemingly a result of the reduction process responsible for separating the system into two temperature regimes as suggested previously (Anke et al., 2019). During the formation of the cracks in the ESEM experiment, it can also be seen in Figure 3 that the cracks mostly form in the middle of the ridges of the catalyst structure, i.e. the parts of the surface in between the holes. This is an irreversible phase change, where the cracks persist throughout the cooling and the second heating cycle of the experiment. It can also be seen from the quadrupole mass spectrometer (QMS) data in Figure 2 that a second heating cycle of the experiment with the same catalyst and the already present cracks does not exhibit the same selectivity toward acetone as the first cycle. Thus, insight into the evolution of the cracks is important in order to establish more realistic structure–function correlations. The goal is to semantically segment this crack feature (assigning a binary classification per pixel). The formation of the crack feature in the image is the manifestation of a change in the catalyst surface, and it appears following a significant change in selectivity toward acetone. With the semantic segmentation data, statistical analysis can be performed, providing a deeper understanding of this process (Treder et al., 2023). For this experimental dataset, trivial image processing methods are not sufficient to isolate the cracks: The holes and the cracks occupy the same color space and have a similar color difference from the background, making techniques like thresholding, or edge detection unsuitable.

Synthetic Dataset for Model Training

A set of backgrounds was manually taken from catalyst surface images before the surface transformation. The set of background images was augmented through random rotation, translation, and scaling. To generate physics-based crack structures, the holes in the surface structure of the background image are first identified by using a threshold function, using an adjusted form of Otsu's method (Otsu, 1979). The threshold value resulting from Otsu's method is multiplied by 0.9 to improve the definition of the holes. A distance transform of the resulting binary hole map is performed. To calculate the distance transform, the distance between each pixel and its nearest hole pixel is computed. The result of this process provides a map of distances to holes, the distance transform map **J**. The synthetic cracks are then generated according to the algorithm illustrated in Figure 4 and superimposed on the corresponding background image. The algorithm takes the distance transform map **J** as an input and outputs a trajectory for a synthetic crack feature that respects the physical characteristics of the real crack features.

The algorithm utilizes an iterative process to generate these trajectories: The starting point of the trajectory is given a random x, y coordinate L_x^0, L_y^0 and angle value L_θ^0 (rad). The position and angle of each following point are then given by the following, where s determines the step size and r is randomly chosen from its associated interval in each step:

$$L_x^n = L_x^{n-1} + s\cos(L_\theta^{n-1}) \quad (1)$$

$$L_y^n = L_y^{n-1} + s\sin(L_\theta^{n-1}) \quad (2)$$

$$L_\theta^n = L_\theta^{n-1} + r, \quad r \in [-0.5, 0.5] \quad (3)$$

These equations describe the movement of the point in the trajectory according to its assigned angle in equations (1) and (2), followed by a limited randomization of its direction in equation (3). After each iteration, the value of the distance transform map **J** at the current position (L_x^n, L_y^n) is checked to be above the threshold of five pixels. If this condition is not met, the trajectory ends. As such, the trajectory of the synthetic cracks cannot come within a radius of five pixels of a hole in the background image, and thus stays on the ridges of the catalyst surface. In Figure 4, this is represented by the fourth circle. Multiple trajectories are generated per background image through repetition of this algorithm. Values such as the threshold and angle change intervals were chosen based on empirical knowledge from the experimental dataset.

Due to the presence of the LSTM element in the neural network architecture, the neural network must be trained with sequential time-series data. The trajectories generated by the previously described algorithm are, therefore, used to generate synthetic data sequences with evolving crack features. This is achieved by adding the values of a parametric kernel around each point of the sequence. The kernel parameters control the intensity and width of the kernel, and thereby the resulting crack feature. The kernel parameters are systematically varied within the sequence to mimic the evolution of the cracks on the surface over time. A temporal parameter τ is varied over the length of the sequence, which influences the kernel parameters, generating thicker and more intense lines for higher values of τ. To generate a diverse synthetic dataset, four different

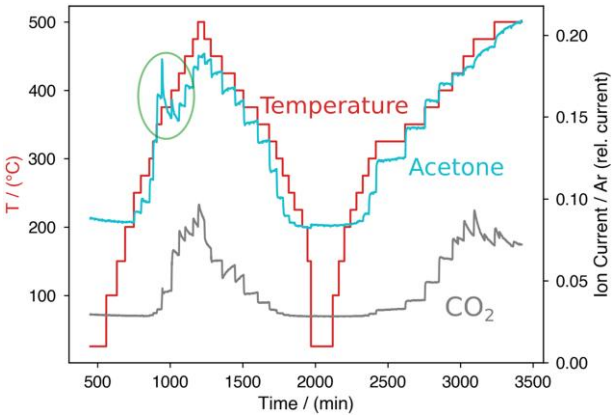

Fig. 2. Selected QMS output of the ESEM experiment. The graph shows the ion current intensities of the desired selective oxidation of isopropanol to acetone, and the undesired full oxidation to carbon dioxide. It can be seen that there is a peak in the acetone production (circled) in the first heating cycle prior to the increase in CO_2 production that is not present in the second cycle.

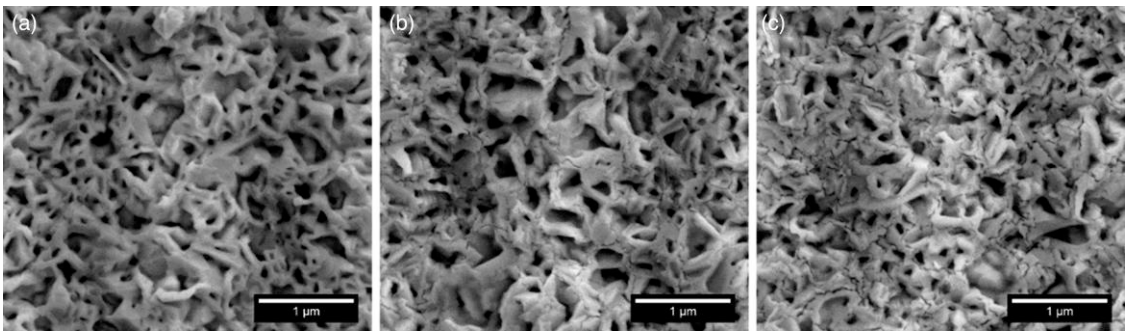

Fig. 3. Experimental images captured with the LFD of the ESEM. Crack formation in the reaction time-series. Images shown are 50 frames apart, with a scan time per frame of 35.40 s. The area captured in the images is different for each image. Contrast has been increased to make features more easily discernible. The important features are the "holes", dark areas visible in (**a,b,c**) and the "cracks", the thin dark lines additionally visible in (**b,c**).

types of sequences are generated as follows, where N is the length of the sequence:

$$\text{Full sequence}: \quad \tau = \{0, 1, 2, \ldots, N\} \tag{4}$$

$$\text{Slow sequence}: \quad \tau = \left\{n, n + \frac{1j}{N}, n + \frac{2j}{N}, \ldots, n + j\right\}$$

$$\text{With}: \quad n \in \{0, \ldots, N\}, \quad j \in \{1, 2, 3\} \tag{5}$$

$$\text{Constant sequence}: \quad \tau = n, \quad n \in \{1, \ldots, N\} \tag{6}$$

$$\text{Blank sequence}: \quad \tau = 0 \tag{7}$$

These ranges describe the different possibilities for the temporal parameter τ. In the full sequence equation (4), the range of τ comprises the full possible range, resulting in a sequence that shows synthetic cracks growing from nothing to full intensity. The slow sequence equation (5) comprises a subset of the full sequence with a higher time resolution. The parameters j, n are randomly chosen from their associated intervals for each sequence.The same sequence equation (6) keeps τ fixed at a randomly chosen n. The empty sequence equation (7) keeps τ fixed at 0, adding the backgrounds with no cracks into the synthetic dataset. This algorithm was implemented using Python, with the OpenCV library (Bradski, 2000).

Finally, a noise layer was added to the image, to approximate the noise in the experimental dataset. This is represented as a random value from a white noise distribution taken for each pixel and added to its value. Representative results of this algorithm can be seen in Figure 5. By construction, the generated cracks avoid the holes in the background, following the distance map. Using this algorithm, a physics-based synthetic dataset was generated and used to train a neural network. To provide a point of comparison, a random synthetic dataset was also generated and used to train a different neural network as described below. The random synthetic dataset was generated using the same algorithm as the physics-based synthetic dataset, with the end condition based on the distance transform of the background omitted. This results in a synthetic dataset where the synthetic crack trajectories are taken from unguided random walkers. The main difference between the two synthetic datasets will thereby be the presence of overlapping holes and cracks within the random synthetic dataset. It should be noted that the generation of the trajectories of the random synthetic dataset is still constructed with the restricted angle movement based on the real cracks and therefore is still partially based on the physics of the real system.

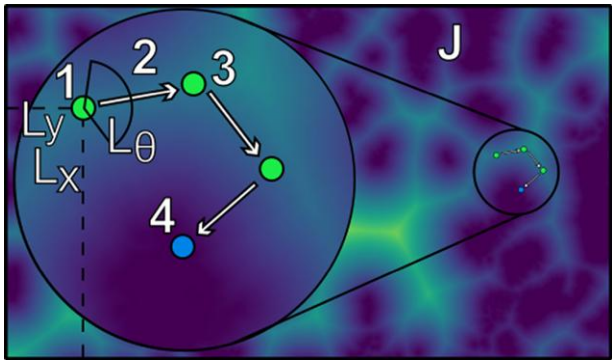

Fig. 4. Synthetic crack generation algorithm. (1) A point is placed with random L_x, L_y coordinates and a random angle L_θ. (2) According to the angle and position of the first point, the location of the next point is determined. (3) The distance transform map value at the new coordinates L_x, L_y is checked to be above the threshold, and the angle L_θ is randomly adjusted. (4) Steps 2 and 3 are repeated until the distance transform map value is below the threshold of five pixels distance from a hole. In this case, the algorithm is ended and all points are saved as a sequence. If an initial point fails the threshold check, the trajectory is discarded, leading to a varying number of trajectories in each sequence. Colors shown are qualitative, with darker pixels indicating holes and brighter pixels indicating areas further away from holes, greater than the threshold.

Neural Network Architecture for Segmentation Model

The network architecture that was used is based on the U-NET architecture as described in the *Introduction*, with ResizeConv (Odena et al., 2016) layers replacing ConvTranspose layers. The ConvTranspose layer used in the original U-NET introduces checkerboard artifacts into the segmentation mask with each upsampling. ResizeConv layers provide an alternative upsampling method: the image is resized, then convolution is performed with a 3×3 trainable kernel. Additionally, a ConvLSTM (Shi et al., 2015) layer was placed after the final layer of the U-NET. The ConvLSTM layer introduces a recurrent element to the network, allowing it to take advantage of the sequential nature of the input data.

The model was implemented and trained using the PyTorch library (Paszke et al., 2019), using cross entropy loss and a

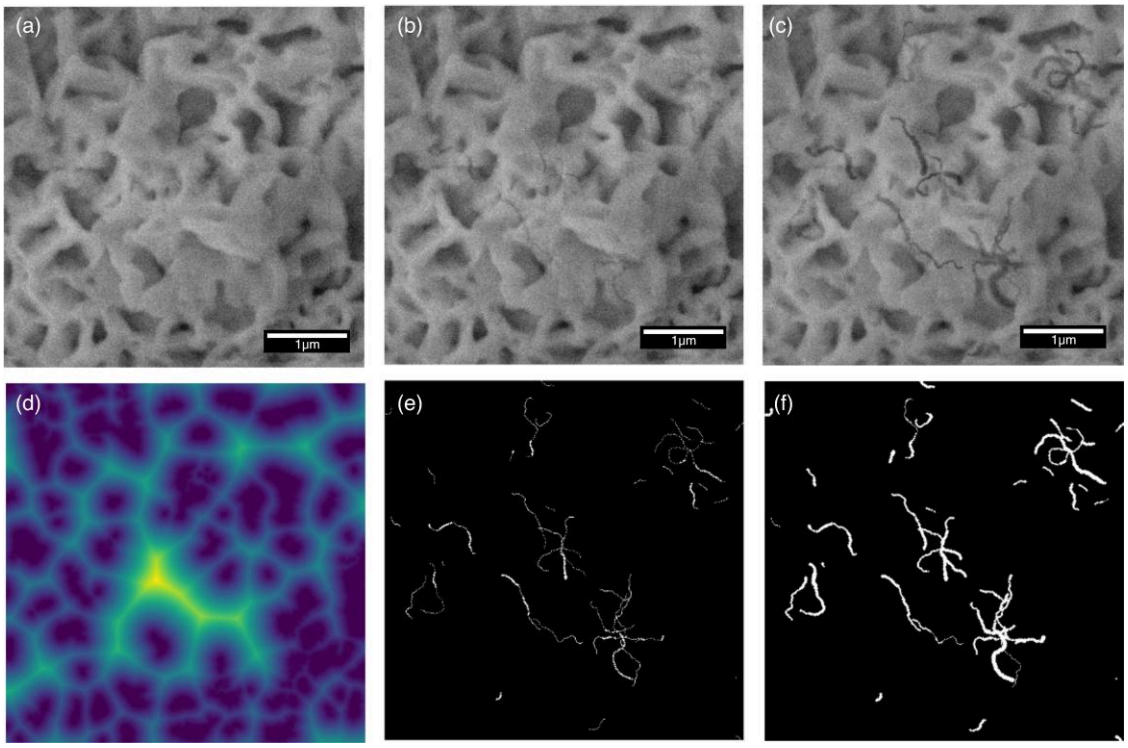

Fig. 5. (**a**) Experimental image captured with the LFD of the ESEM. Background taken from an initial room-temperature catalyst image before the heating cycle and thus without cracks present. (**b,c**) Synthetic cracks overlaid on experimental image. Examples of synthetic crack evolution superimposed on the background image: Temporal parameter at $\tau = 7$, $\tau = 15$. (**d**) Distance transform: Colors shown are qualitative, with darker pixels indicating holes and brighter pixels indicating areas further away from holes in (**a**). (**e,f**) Synthetic-generated masks, used for training data as ground truth, corresponding to (**b, c**). Generated cracks only occupy bright areas of the distance transform map, showcasing the effect of the distance transform map threshold.

stochastic gradient descent optimizer with an unoptimized learning rate of 0.8. The final synthetic dataset consisted of 768 synthetic training sequences consisting of 16 512×512 images per sequence, and 256 synthetic validation sequences, with a minibatch size of eight sequences. These parameters were chosen through manual hyperparameter optimization. The model was trained on an NVIDIA A100 80GB GPU for 70 epochs. The optimal epoch was then chosen by minimizing the validation and training error, and by evaluating the performance on the experimental dataset. The implementation of the neural network and synthetic data generator can be found in Supplementary Material S2.

Results

Model Performance Benchmark

To show the impact of the physics-based distance map component of the synthetic dataset described in the *Dataset* section, a benchmark was performed between two networks with an identical architecture (described in the *Synthetic Dataset for Model Training* section). The physics-based network was trained with the physics-based synthetic dataset. The random network was trained with the random synthetic dataset.

The benchmark consists of a challenging synthetic test dataset. To generate the synthetic test dataset, the same algorithm was used as for the physics-based synthetic dataset, with fewer lines (five lines), generated per image. Furthermore, temporal information was not supplied in this case: each frame in the sequence was generated with the same iteration value τ, leading to sequences with differences only in noise. In this process, an individual mask was stored for each synthetic crack. The binarized output of the models was then evaluated as follows: First, to check for true positives (TP), a threshold of pixels marked as cracks by the model within each individual mask must be reached. Any sufficiently large cluster of marked pixels that is not within any of the masks is then marked as a false positive (FP), and any unmarked pixels within the masks are marked as false negative (FN). Any remaining marked pixels that were not assigned to the other categories are counted as noise pixels. A pixel-wise Dice score (Dice, 1945) is calculated as

$$\frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{8}$$

where the FP value includes both clustered false positive pixels and noise pixels (Carass et al., 2020).

It was found that much of the error in the pixel-wise evaluations originated from the model correctly identifying the trajectory of a specific crack but differing from the ground truth in the thickness of the crack, resulting in false positive or false negative errors. In order to evaluate the performance of the model purely in identifying the trajectories of the cracks, the metric was changed from a pixel-wise evaluation to the number of cracks (out of the five present) identified using this cluster approach.

The results of this benchmark can be seen Figure 6. It is clear that the physics-based model, in comparison to the random walk model, exhibits an overall decrease in the amount of false positives at the cost of a minor penalty to the amount of true positives when evaluating the synthetic test dataset. The Gaussian shapes of the distributions of false positives in Figure 6 are expected (following the central limit theorem), and show that the synthetic test dataset generated for the

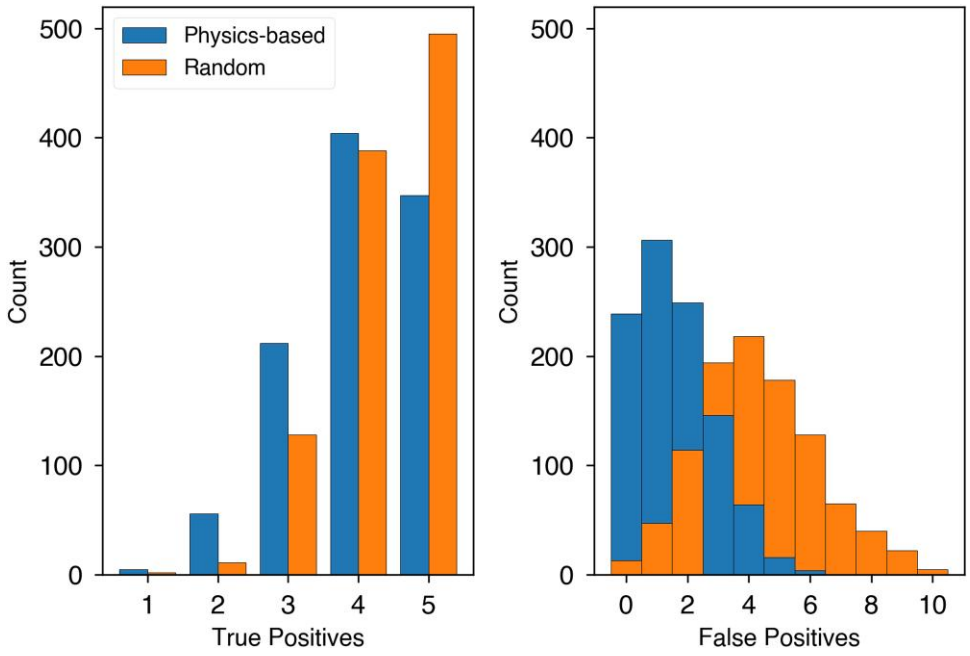

Fig. 6. Accuracy distribution of synthetic data benchmark. Data from 1,024 synthetic test dataset sequences. Input data consisted of sequences with five generated crack features persistently present across frames in each sequence. As such, an ideal segmentation would be five true positives and no false positives. Segmented and classified as described by two neural networks, one trained with a physics-based synthetic dataset, the other trained with a random synthetic dataset. The resulting Dice scores are 0.76 (physics-based) and 0.64 (random).

benchmark contains a reliable and diverse distribution of challenging segmentation tasks for the model. The effective result of the physics-based component included in the physics-based synthetic dataset is that the model becomes much more specific toward the crack feature, rather than the holes. This is in accordance with the expectation, as the random synthetic dataset frequently includes the situation where a synthetic crack feature overlaps with a hole, confusing the model's trained filters. Despite coming at a small cost to the rate of recall of the feature, this change still infers an improvement of the Dice score from the reduction in false positives.

ESEM Data Time-Series Segmentation

In this section, we report the results obtained by applying the neural network trained with the physics-based synthetic dataset (as described in the *Materials and Methods* section) to the ESEM experimental dataset.

The model was applied to the whole experimental dataset of the ESEM time-series. The result of this can be found as a video in Supplementary Material S3, as well as an example frame in Figure 7. It can be seen that the model provides a reliable segmentation of the dataset. The model is able to segment the crack features as soon as they begin forming on the surface, without any false positives prior to the formation of the first crack feature. The segmentation remains stable through the sudden scene changes and drifting within the microscopy data resulting from changes in temperature in the experiment. For the intended application of the detection of the formation of the crack feature on the surface, the reduction of false positives is of great importance. The observed reduction in the area under the curve of the physics-based segmentation output, when compared to the random segmentation output (as illustrated in Fig. 8), can be attributed to such a decrease in false positives. This suggests that the physics-based model provides segmentation results that are more aligned with the ground truth. Notably, within the low-temperature regime (prior to crack formation, occurring before 300 min), the physics-based component significantly reduces false positives, highlighting the role of the physics-based component in the ability of the model to accurately identify the initial nucleation point of the crack feature.

Discussion

Comparing the segmented output of the model (crack amount) in Figure 8 to the selectivity of the catalyst over the experiment, we can get a full picture of the relation between these factors. From this, it can be seen that the crack formation occurred on the surface ~100 min after the maximum in acetone selectivity ceased, and that a continuously increasing temperature is required to facilitate further crack growth. As the relatively low resolution of the ESEM does not show us the full nanoscale image of the surface, it appears to us that the cracks are a manifestation of an overall surface phase change that is initiated around 350°C and continues to affect the surface until the cracks are fully formed. We intend to investigate the nucleation of the cracks further, by letting the trained neural network control the microscope. Upon detection of the onset of crack formation, the neural network would send a command to the microscope to immediately quench the reaction, stopping the heating and gas flow. Then, the catalyst can be placed in a higher-resolution electron microscopy setup such as immersion-SEM or TEM to gain insights into the nature of the phase change (Kaegi & Holzer, 2003; Cao et al., 2020; Sandoval-Diaz et al., 2020; Govind et al., 2024).

A restriction of the physics-based synthetic data model approach is the fact that it must be possible to approximately model the physical properties of the system. Some systems

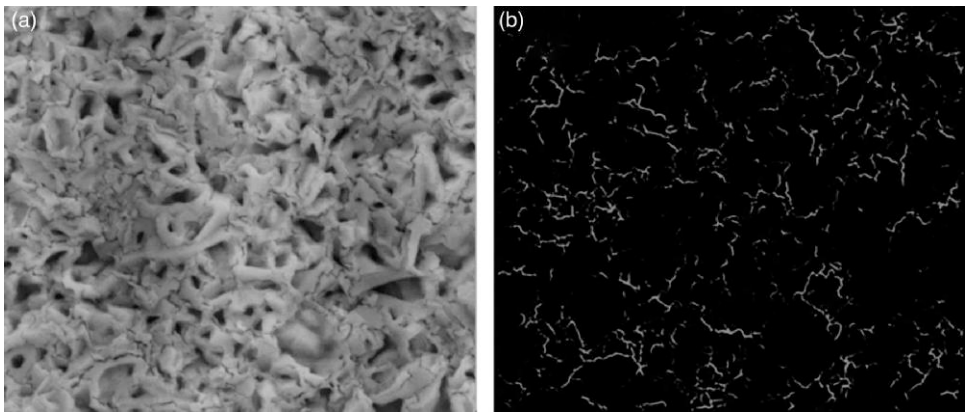

Fig. 7. Example segmentation frame, taken from the full segmentation video. (**a**) Experimental image captured with the LFD of the ESEM. Input image from experimental dataset showing cracks in the surface. Contrast has been enhanced for clarity (Scheurer et al., 2024). (**b**) Segmentation output of the machine learning model trained with physics-based synthetic data, with (**a**) as input. Brightness indicates confidence in the presence of a crack.

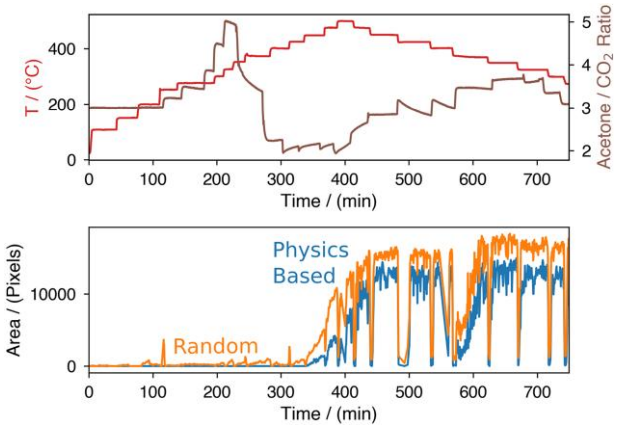

Fig. 8. Segmentation output of the physics-based and random models over the first 800 min of the ESEM experimental dataset. Area is taken as a sum of activated pixels of the model output. Spiked dips in the output are the result of image shifting due to temperature changes.

may be too complex or otherwise infeasible to apply this approach to. Furthermore, a diverse enough set of background (areas with no features) images is a prerequisite to apply this method. We plan to investigate the integration of theoretical simulation techniques into the construction of synthetic data models. Data obtained from such simulations could be used directly in the construction of synthetic data models (through rendering of structures into images) or supplement known experimental results.

It is likely that the segmentation performance of the physics-based synthetic data model could be further improved through experimentation with hyperparameters of the U-NET network architecture, entirely different network architectures and other machine learning techniques such as utilizing pre-trained networks (Hinterstoisser et al., 2017). However, as the focus of this work was on the development of the physics-based synthetic data generation method, this was considered out of scope for this work.

Summary

In this paper, we have shown that the segmentation of a feature on a complex catalyst surface can be accurately performed by a neural network trained from random initial parameters only with synthetic data. By integrating physical characteristics of the features on the catalyst surface into the synthetic training data generation algorithm, we achieved superior accuracy and greatly decreased false positives.

Availability of Data and Materials

Source code and data used in this article can be found at Scheurer et al. (2024).

Supplementary Material

To view supplementary material for this article, please visit http://academic.oup.com/mam/article-lookup/doi/10.1093/mam/ozae130#supplementary-data.

Acknowledgments

We thank Dr. Simone Köcher [Institute of Energy Technologies, Fundamentals of Electrochemistry (IET-1), Forschungszentrum Jülich GmbH, Jülich, Germany] for proofreading and correcting this work.

Financial Support

We acknowledge funding from the German Federal Ministry of Education and Research in the framework of the project CatLab (03EW0015A). This project was in part funded by the Deutsche Forschungsgemein-schaft (DFG, German Research Foundation)—388390466-TRR 247, subproject B6. This work was further supported by the DFG under Germany's Excellence Strategy—EXC 2089/1—390776260.

Conflict of Interest

The authors declare that they have no competing interest.

References

Anke S, Bendt G, Sinev I, Hajiyani H, Antoni H, Zegkinoglou I, Jeon H, Pentcheva R, Roldan Cuenya B, Schulz S & Muhler M (2019). Selective 2-propanol oxidation over unsupported Co_3O_4 spinel nanoparticles: Mechanistic insights into aerobic oxidation of alcohols. *ACS Catal* **9**(7), 5974–5985. https://doi.org/10.1021/acscatal.9b01048

Bermudez-Chacon R, Marquez-Neila P, Salzmann M & Fua P (2018). A domain-adaptive two-stream U-Net for electron microscopy image

segmentation. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE. https://doi.org/10.1109/isbi.2018.8363602

Bradski G (2000). The OpenCV Library. Dr. Dobb's Journal of Software Tools, Version 4.8.1. https://github.com/opencv/opencv/releases

Cao J, Rinaldi A, Plodinec M, Huang X, Willinger E, Hammud A, Hieke S, Beeg S, Gregoratti L & Colbea C (2020). In situ observation of oscillatory redox dynamics of copper. *Nat Commun* **11**, 3554. https://doi.org/10.1038/s41467-020-17346-7

Carass A, Roy S, Gherman A, Reinhold JC, Jesson A, Arbel T, Maier O, Handels H, Ghafoorian M, Platel B, Birenbaum A, Greenspan H, Pham DL, Crainiceanu CM, Calabresi PA, Prince JL, Gray Roncal WR, Shinohara RT & Oguz I (2020). Evaluating white matter lesion segmentations with refined Sørensen-Dice analysis. *Sci Rep* **10**(1). https://doi.org/10.1038/s41598-020-64803-w

Chang WS & Liao SW (2019). KUnet: Microscopy image segmentation with deep Unet based convolutional networks. In *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*. IEEE. https://doi.org/10.1109/smc.2019.8914048

Danilatos GD (1983). A gaseous detector device for an environmental SEM. *Micron Microsc Acta* **14**(4), 307–318. https://doi.org/10.1016/0047-7206(83)90002-X

DeCost BL, Lei B, Francis T & Holm EA (2019). High throughput quantitative metallography for complex microstructures using deep learning: A case study in ultrahigh carbon steel. *Microsc Microanal* **25**(1), 21–29. https://doi.org/10.1017/S1431927618015635

Dice LR (1945). Measures of the amount of ecologic association between species. *Ecology* **26**(3), 297–302. https://doi.org/10.2307/1932409

Donald AM (2003). The use of environmental scanning electron microscopy for imaging wet and insulating materials. *Nat Mater* **2**(8), 511–516. https://doi.org/10.1038/nmat898

Ede JM (2021). Deep learning in electron microscopy. *Mach Learn Sci Technol* **2**(1), 011004. https://doi.org/10.1088/2632-2153/abd614

Fawaz HI, Forestier G, Weber J, Idoumghar L & Muller P-A (2018). 'Data augmentation using synthetic data for time series classification with deep residual networks.' arXiv, arXiv:1808.02455v1, preprint: not peer reviewed. https://doi.org/10.48550/arXiv.1808.02455

Govind K, Oliveros D, Dlouhy A, Legros M & Sandfeld S (2024). Deep learning of crystalline defects from TEM images: A solution for the problem of 'never enough training data'. *Mach Learn Sci Technol* **5**(1), 015006. https://doi.org/10.1088/2632-2153/ad1a4e

Groschner CK, Choi C & Scott MC (2021). Machine learning pipeline for segmentation and defect identification from high-resolution transmission electron microscopy data. *Microsc Microanal* **27**(3), 549–556. https://doi.org/10.1017/S1431927621000386

Hinterstoisser S, Lepetit V, Wohlhart P & Konolige K (2017). 'On pre-trained image features and synthetic images for deep learning.' arXiv, arXiv:1710.10710v2, preprint: not peer reviewed. https://doi.org/10.48550/arXiv.1710.10710

Hochreiter S & Schmidhuber J (1997). Long short-term memory. *Neural Comput* **9**(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Kaegi R & Holzer L (2003). Transfer of a single particle for combined ESEM and TEM analyses. *Atmos Environ* **37**(31), 4353–4359. https://doi.org/10.1016/S1352-2310(03)00574-0

Liu J, Hong B, Chen X, Xie Q, Tang Y & Han H (2021). An effective AI integrated system for neuron tracing on anisotropic electron microscopy volume. *Biomed Signal Process Control* **69**, 102829. https://doi.org/10.1016/j.bspc.2021.102829

Madsen J, Liu P, Kling J, Wagner JB, Hansen TW, Winther O & Schiøtz J (2018). A deep learning approach to identify local structures in atomic-resolution transmission electron microscopy images. *Adv Theory Simul* **1**(8). https://doi.org/10.1002/adts.201800037

Odena A, Dumoulin V & Olah C (2016). Deconvolution and checkerboard artifacts. *Distill* https://doi.org/10.23915/distill.00003.

Otsu N (1979). A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern* **9**(1), 62–66. https://doi.org/10.1109/TSMC.1979.4310076

Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, Desmaison A, Kopf A, Yang E, DeVito Z, Raison M, Tejani A, Chilamkurthy S, Steiner B, Fang L, Bai J & Chintala S (2019). Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pp. 8024–8035. Curran Associates, Inc.

Ronneberger O, Fischer P & Brox T (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv* **arXiv:1505.04597v1**, preprint: not peer reviewed. https://doi.org/10.48550/arXiv.1505.04597

Sandoval-Diaz L, Cruz D, Vuijk M, Ducci G, Hävecker M, Jiang W, Plodinec M, Hammud A, Ivanov D, Götsch T, Reuter K, Schlögl R, Scheurer C, Knop-Gericke A & Lunkenbein T (2024). Metastable nickel-oxygen species modulate rate oscillations during dry reforming of methane. *Nat Catal* **7**(2), 161–171. https://doi.org/10.1038/s41929-023-01090-4

Sandoval-Diaz L, Plodinec M, Ivanov D, Poitel S, Hammud A, Nerl HC, Schlögl R & Lunkenbein T (2020). Visualizing the importance of oxide-metal phase transitions in the production of synthesis gas over Ni catalysts. *J Energy Chem* **50**, 178–186. https://doi.org/10.1016/j.jechem.2020.03.013

Scheurer C, Vuijk M, Ducci G, Sandoval L, Pietsch M, Reuter K & Lunkenbein T (2024). Physics-based synthetic data model for automated segmentation in catalysis microscopy; data and source code. https://doi.org/10.17617/3.NWOKER

Schneider NM, Park JH, Norton MM, Ross FM & Bau HH (2016). Automated analysis of evolving interfaces during in situ electron microscopy. *Adv Struct Chem Imaging* **2**, 2. https://doi.org/10.1186/s40679-016-0016-z

Shah A, Schiller JA, Ramos I, Serrano J, Adams DK, Tawfick S & Ertekin E (2023). Automated image segmentation of scanning electron microscopy images of graphene using U-Net neural network. *Mater Today Commun* **35**, 106127. https://doi.org/10.1016/j.mtcomm.2023.106127

Shi X, Chen Z, Wang H, Yeung D-Y, Wong W-K & Woo W-C (2015). 'Convolutional LSTM network: A machine learning approach for precipitation nowcasting.' arXiv, arXiv:1506.04214v1, preprint: not peer reviewed. https://doi.org/10.48550/arXiv.1506.04214

Toth P, Farrer JK, Palotas AB, Lighty JS & Eddings EG (2013). Automated analysis of heterogeneous carbon nanostructures by high-resolution electron microscopy and on-line image processing. *Ultramicroscopy* **129**, 53– 62. https://doi.org/10.1016/j.ultramic.2013.02.017

Trampert P, Rubinstein D, Boughorbel F, Schlinkmann C, Luschkova M, Slusallek P, Dahmen T & Sandfeld S (2021). Deep neural networks for analysis of microscopy images-synthetic data generation and adaptive sampling. *Crystals (Basel)* **11**(3), 258. https://doi.org/10.3390/cryst11030258

Treder KP, Huang C, Bell CG, Slater TJA, Schuster ME, Özkaya D, Kim JS & Kirkland AI (2023). nNPipe: A neural network pipeline for automated analysis of morphologically diverse catalyst systems. *NPJ Comput Mater* **9**, 18. https://doi.org/10.1038/s41524-022-00949-7.

Uwins PJR, Millar GJ & Nelson ML (1997). Dynamic imaging of structural changes in silver catalysts by environmental scanning electron microscopy. *Microsc Res Tech* **36**(5), 382. https://doi.org/10.1002/(ISSN)1097-0029

Wilms M, Handels H & Ehrhardt J (2017). Multi-resolution multi-object statistical shape models based on the locality assumption. *Med Image Anal* **38**, 17–29. https://doi.org/10.1016/j.media.2017.02.003

Ziatdinov M, Dyck O, Maksov A, Li X, Sang X, Xiao K, Unocic RR, Vasudevan R, Jesse S & Kalinin SV (2017). Deep learning of atomically resolved scanning transmission electron microscopy images: Chemical identification and tracking local transformations. *ACS Nano* **11**(12), 12742–12752. https://doi.org/10.1021/acsnano.7b07504